

Children's Vocal Analysis from Naturalistic Recordings Based Artificial Neural Network for Autism Detection

Royan Dawud Aldian^{1,a}, Endah Purwanti^{2,b}, and Soegianto Soelistono^{3,c}

¹ Biomedical Engineering, Faculty of Science and Technology, Universitas Airlangga, Indonesia.

Email: royandawud-09@fst.unair.ac.id

^{2,3} Departement of Physic, Faculty of Science and Technology, Universitas Airlangga, Indonesia

^b Email: endah-p-1@fst.unair.ac.id

^c Email: soegianto@fst.unair.ac.id

Abstract—In this research we have been developed an automatic investigation to classify normal children voice or autistic by using modern computation technology that is computation based on artificial neural network. The superiority of this computation technology is its capability on processing and saving data. In this research, digital voice features are gotten from the coefficient of linier predictive coding with autocorrelation method and have been transformed in frequency domain using fast fourier transform, which used as input of artificial neural network in backpropagation method so that will make the difference between normal children and autistic automatically. The result of backpropagation method shows that successful classification capability for normal children voice experiment data is 100% whereas, for autistic children voice experiment data is 100%. The success rate using backpropagation classification system for the entire test data is 100%.

Keywords—autism; artificial neural network; backpropagation; linier predictive coding; fast fourier transform.

I. INTRODUCTION

According to data from UNESCO in 2011, there were 35 million people with autism around the world. The average, 6 of 1000 people in the world have suffered from autism. In the United States, autism is owned by 11 of the 1000 people. While in Indonesia, comparison is 8 of 1000 people. This figure is quite high considering the counted in 1989, only two people known to have autism [1].

Although the experts claimed for years that children with autism have a different voice when speaking compared to normal children, there is no practical way to use vocalization as a part of the diagnostic or screening process in line with autism [2].

Nowadays, the improvement of technology in computation field is developed rapidly. Computation technology can be used in processing and saving data. Artificial neural network is one of computation method which has many over plus, such as its capability in doing prediction with non-linier system, speedy finishing time and robust for missing data. In this research, backpropagation method based artificial neural network is used to classify the children's voices, so that it can be distinguished normal or autistic children automatically.

II. LITERATURE REVIEWS

A. Autism

So far have not been discovered yet a clinical test that can diagnose autism directly. To establish a diagnosis of autism disorders clinicians often use the DSM IV guidelines. Autism disorders are diagnosed based on based on the instructions in the DSM-IV. The best diagnosis is by carefully observing the

behavior of children in communicating, behavior and level of development [3].

Autism disorders begin in childhood. Infantile autism (autism in childhood) is Pervasive Developmental Disorders (PDD) which includes the inability to interact with others, the language disorder is shown with delayed proficiency, echolalia, mutism (silence, do not have the ability to speak), and the reversal of the sentence. [4].

B. The Signal Sampling Process

Sampling was done at a speed of 8000 Hz with a resolution of 8 bits (1 byte) in order to get as much data as 8000 bytes per second. Speed of sampling was done with the assumption that the speech signal is in the region 300-3400 Hz frequency that meets the Nyquist criteria states [5]:

$$f_s \geq 2f_h f_h = f_{intop} \quad (1)$$

C. Linier Predictive Coding

An incoming speech signal is segmented or framed with a certain frame length. Then the framed speech was carried out with windowing which can be done by using hamming window, which has the equation [6]:

$$w(n) = 0,54 - 0,46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \leq n \leq N-1 \quad (2)$$

where $n = 0, 1, \dots, M-1$, and M is the length of the frame, so we get a framed speech signals that have been windowed by using hamming window with the sequence:

$$s_n(m) = s(n+m) \cdot w(n), \quad 0 \leq m \leq N-1 \quad (3)$$

LPC analysis was done by using the autocorrelation method, it assumes that the signal has a value equal to zero for intervals outside the analyzed region ($0 \leq n \leq N-1$). The following equations form the autocorrelation function:

$$R_n(i-k) = \sum_{m=0}^{N-1-(i-k)} s_n(m) s_n(m+i-k) \quad (4)$$

If the autocorrelation function is symmetric $R_n(k) = R_n(-k)$, so the LPC equations can be denoted as:

$$\sum_{k=1}^p R_n(|i-k|) a_k = R_n(i), \quad 1 \leq i \leq p \quad (5)$$

LPC coefficients values obtained by solving the following matrix equation:

$$\begin{bmatrix} R_n(0) & R_n(1) & R_n(2) & \dots & R_n(p-1) \\ R_n(1) & R_n(0) & R_n(1) & \dots & R_n(p-2) \\ R_n(2) & R_n(1) & R_n(0) & \dots & R_n(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ R_n(p-1) & R_n(p-2) & R_n(p-3) & \dots & R_n(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \dots \\ a_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ R_n(3) \\ \dots \\ R_n(p) \end{bmatrix} \quad (6)$$

D. Fast Fourier Transform

Fast Fourier Transform is a simplification of the Discrete Fourier Transform (DFT). For discrete time signal $x(n)$, the DFT of the signal is given by [7]:

$$H(k) = \sum_{n=0}^{N-1} h(n) e^{-\frac{j2\pi nk}{N}}, \text{ untuk } 0 \leq k \leq (N-1) \quad (7)$$

Algorithms Fast Fourier Transform (FFT) is an efficient DFT calculation procedure that will accelerate the process of calculating the DFT. When applied to the area of time then the algorithm is referred to as FFT decimation in time (DIT). Decimation then leads to a significant reduction in the number of calculations performed on the data area of the time.

Equation (7) becomes:

$$H_1(k) = \sum_{n=0}^{N/2-1} h(n)W_N^{nk}, \text{ untuk } 0 \leq k \leq (N-1) \quad (8)$$

where the factor $e^{-\frac{j2\pi nk}{N}}$ will be written as:

$$W_N^{nk} = e^{-\frac{j2\pi nk}{N}} = \cos(2\pi/N) - j \sin(2\pi/N) \quad (9)$$

Suffix n in Equation (8) extended from $n = 0$ to $n = N-1$, corresponding to a data value of $h(0), h(1), h(2), h(3) \dots h(N-1)$. Even-numbered sequence is $h(0), h(2), h(4) \dots h(N-2)$ and the odd-numbered sequence is $h(1), h(3) \dots h(N-1)$. The second sequence contains $N/2$ -titik. Sequence may even be marked $h(2n)$ with $n = 0$ to $n = N/2-1$, while the odd sequence into $h(2n-1)$. Then Equation (8) can be rewritten as:

$$H_1(k) = \sum_{n=0}^{N/2-1} h(2n)W_N^{2nk} + \sum_{n=0}^{N/2-1} h(2n+1)W_N^{2(2n+1)k} \\ = \sum_{n=0}^{N/2-1} h(2n)W_N^{2nk} + W_N^k \sum_{n=0}^{N/2-1} h(2n+1)W_N^{2nk}, 0 \leq k \leq (N-1) \quad (10)$$

Furthermore, by replacing $W_N^{2nk} = W_{N/2}^{nk}$, then Equation (10) becomes:

$$H(k) = \sum_{n=0}^{N/2-1} h(2n)W_{N/2}^{nk} + W_{N/2}^k \sum_{n=0}^{N/2-1} h(2n+1)W_{N/2}^{nk} \quad (11)$$

E. Artificial Neural Network

ANN is an artificial representation of the human brain that tries to simulate the learning process of the human brain. ANN consists of some nodes that is processor element. Every node presenting a neuron. The relations between nodes are reached with connection weight. The changing that occurs during the learning process is the change in the value of the weight.

One method to perform supervised competitive layer learning is backpropagation. Back propagation is a systematic method for training multilayer neural networks. This method has a solid mathematical foundation, and the objective is to get the form of equation algorithm and the coefficient in the formula by minimizing the error sum of squares error through the model developed [8].

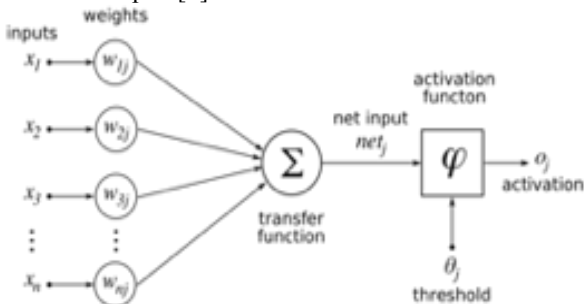


Fig 1. Model of Artificial Neuron

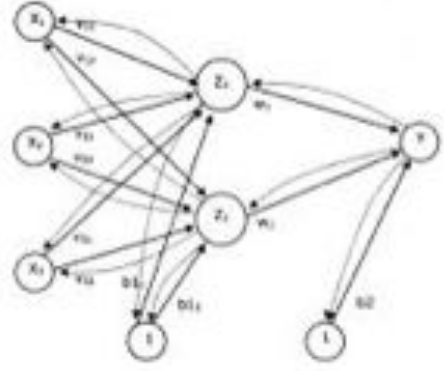


Fig 2. BackpropagationArchitecture [8]

III. METHODS

The data of this research were 26 children voice recording consisting of 19 normal children voices recording data and 7 autistic children voices recording data. Data which is used for ANN training were 20 data. While data which is used for ANN testing were 6 data. All data taken from normal and autistic children who have the age in the range of 3-7 years. Classification of children's voices used ANN with Backpropagation method. Research flow chart shown in Fig. 3.

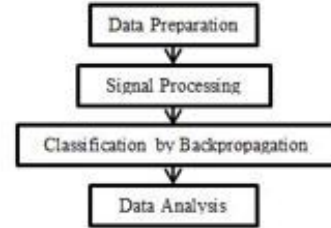


Fig 3. Flow chart of research

Then the results of the recording are read to obtain a discrete form of speech signal so ready for a signal processing. Voice signal processing stage aims to extract the signal from voice recording data to obtain extracted signal features that used as the neural network input.

Voice recording readings followed by analysis using the LPC. LPC analysis results in the form of LPC coefficients were made into the template using the autocorrelation method, in order to obtain a representation of the coefficients. Signal processing sequence to obtain the representation values of LPC coefficients using autocorrelation method can be seen in Fig. 4.

LPC analysis begins by determining the values of the parameters needed in the extraction process, such as the order of the signal which has the symbol p, then the length of the frame that has the symbol t, the length of the frame will determine the number of segments obtained from solving the initial signal. The next process is skipped each segment obtained by using Hamming Window and then the LPC coefficients obtained by using the method of autocorrelation in each segment, if the LPC coefficients of all segments have been obtained, then the coefficients stored in the codebook-shape codebook to be used later in the process.

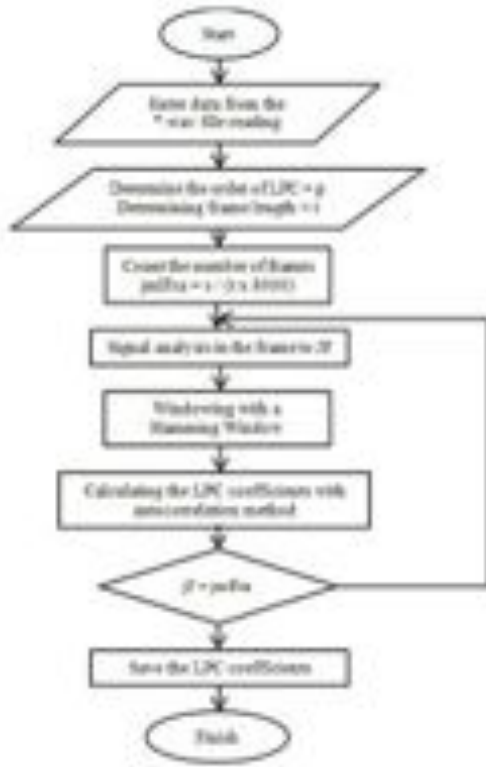


Fig 4. Flow of Voice Signal Extraction Using LPC with Autocorrelation Method

Then all codebook of LPC coefficients will be transformed into the frequency domain using the Fast Fourier Transform (FFT). All data get sampling with the same points, which is 512 in FFT process to obtain matrix feature set sized 512x1 for each voice recording file. Data input for backpropagation input is obtained after matrix feature set of the FFT of each data record was formed.

Classification of voice recordings performed using backpropagation method. The algorithm is as follows:

- Initialize beginning weights with a random value.
- Set :
Maximum epoch, target (t), learning rate (α).
- Initialization:
Epoch = 0; MSE = 1
- Do the following steps for:
Epoch < Maximum Epoch; MSE > Error Target
- 1. Epoch = Epoch + 1
- 2. For each pair of elements that will get the process of learning, do:

Feed forward:

- a. Each input unit (X_i , $i=1,2,3,\dots,n$) receives a signal x_i and forward the signal to the layer that is on it (the hidden layer).
- b. Each unit in the hidden layer (Z_j , $j=1,2,3,\dots,p$) summing the weighted input signals:

$$z_{in_j} = b1_j + \sum_{i=1}^n x_i v_{ij} \quad (12)$$

with the activation function to compute its output signal:

$$z_j = f(z_{in_j}) \quad (13)$$

and the signals are transmitted to all units in the layer above (output units).

- c. Each output unit (Y_k , $k = 1,2,3, \dots, m$) summing the weighted input signals.

$$y_{in_k} = b2_k + \sum_{i=1}^p z_i w_{ik} \quad (14)$$

with the activation function to compute its output signal:

$$y = f(y_{in_k}) \quad (15)$$

and the signals are transmitted to all units in the layer above (output units).

Backpropagation:

- d. Each output unit (Y_k , $k = 1,2,3, \dots, m$) receives a target pattern corresponding to the input pattern learning then calculated the error information:

$$\delta 2_k = (t_k - y_k) f'(y_{in_k}) \quad (16)$$

$$\phi 2_{jk} = \delta 2_k z_j \quad (17)$$

$$\beta 2_k = \delta 2_k \quad (18)$$

then calculated weights and bias correction that will be used to improve the value of w_{jk} and $b2_k$:

$$\Delta w_{jk} = \alpha \phi 2_{jk} \quad (19)$$

$$\Delta b2_k = \alpha \beta 2_k \quad (20)$$

- e. Each hidden unit (Z_j , $j=1,2,3,\dots,p$) summing delta input of the units that are in the layer above it:

$$\delta_{in_j} = \sum_{k=1}^m \delta 2_k w_{jk} \quad (21)$$

that value is multiplied by the derivative of the activation function to calculate the error information:

$$\delta 1_j = \delta_{in_j} f'(z_{in_j}) \quad (22)$$

$$\phi 1_{ij} = \delta 1_j x_i \quad (23)$$

$$\beta 1_j = \delta 1_j \quad (24)$$

then calculated weights and bias correction that will be used to improve the value of v_{ij} dan $b1_j$:

$$\Delta v_{ij} = \alpha \phi 1_{ij} \quad (25)$$

$$\Delta b1_j = \alpha \beta 1_j \quad (26)$$

- f. Each output unit (Y_k , $k=1,2,3,\dots,m$) repair the bias and the weights ($j=0,1,2,\dots,p$):

$$w_{jk}(\text{new}) = w_{jk}(\text{old}) + \Delta w_{jk} \quad (27)$$

$$b2_k(\text{new}) = b2_k(\text{old}) + \Delta b2_k \quad (28)$$

- g. Each hidden unit (Z_k , $k=1,2,3,\dots,p$) repair the bias and the weights ($i=0,1,2,\dots,n$):

$$v_{ij}(\text{new}) = v_{ij}(\text{old}) + \Delta v_{ij} \quad (29)$$

$$b1_j(\text{new}) = b1_j(\text{old}) + \Delta b1_j \quad (30)$$

- 3. Calculate Mean Squared Error (MSE)

After going through the training phase and the final weights obtained (w_{jk}), the next step is testing process with algorithm:

- 1. Initialization:
weight (w_{jk}); bias ($b2_k$)
- 2. Enter the extracted voice recording data to be tested (x_i , $i=1,2,3,\dots,n$)
- 3. Set target (t),
- 4. Do for $i = 1$ to n
 - a. Each input unit (X_i , $i = 1,2,3, \dots, n$) receives the signal x_i and forward the signal to the layer above (hidden layer).
 - b. Each unit in a hidden layer (Z_j , $j = 1,2,3, \dots, p$) summing the weighted input signals:

$$z_{in_j} = b1_j + \sum_{i=1}^n x_i v_{ij} \quad (31)$$

with the activation function to compute its output signal:

$$z_j = f(z_{in_j}) \quad (32)$$

and the signals are transmitted to all units in the layer above (output units).

- c. Each output unit (Y_k , $k = 1, 2, 3, \dots, m$) summing the weighted input signals.

$$y_{in_k} = b2_k + \sum_{i=1}^p z_i w_{jk} \quad (33)$$

with the activation function to compute its output signal:

$$y = f(y_{in_k}) \quad (34)$$

and the signals are transmitted to all units in the layer above (output units).

- d. Output (y) is a classification of data values (x)

IV. RESULTS

The data used in this research is voice recordings that consist of 2 types of voice: normal and autistic children. The data is displayed on a plot that shows the image signal sounds normal children (e.g. Fig. 5a) and children with autism (e.g. Fig. 5b). Each voice signal will enter into the process of windowing with a Hamming Window. Results of signal windowing normal children (e.g. Fig. 6.a) and children with autism (e.g. Fig. 6.b). Results of windowing signal then goes into the process by using the LPC autocorrelation method on each segment of the existing LPC coefficients so obtained for normal children (e.g. Fig. 7.a) and children with autism (e.g. Fig. 7.b). Then Fast Fourier Transform (FFT) is performed after the LPC coefficients obtained as many as 512 data points each, which is normal for a child (e.g. Fig. 8a) and children with autism (e.g. Fig. 8b). FFT is done to improve the performance of software for word patterns to distinguish between each other with the word patterns more clearly.

Feature extraction at each voice recording signal used the LPC with autocorrelation method and continued by FFT produces 512 input parameters for backpropagation method. To obtain optimal parameters of the backpropagation done some variation of the number of neurons in hidden layer, epochs and learning rate. During the training process, the best training is obtained by variation of the number of neurons in hidden layer were 5, 300 epochs and learning rate of 0.1 for faster ANN achieve stability at points 100% and have the smallest learning stoppage than other variations.

Voice recordings for this backpropagation network testing were amounted to 6 data. The data consisted of 4 normal children voices and 2 autistic children voices. Backpropagation network testing results show that the success of the classification ability for normal data testing was 100%, while for autistic data testing was 100%. The success rate using backpropagation classification system for the entire data testing was 100%.

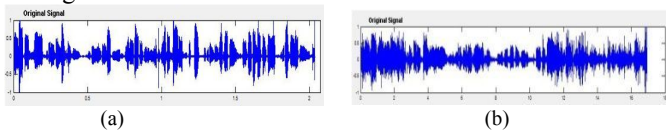


Fig. 5. Original signal of Sample Children Speech

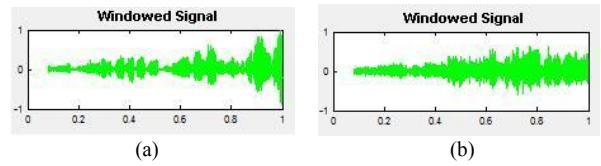


Fig. 6. Windowed Signal

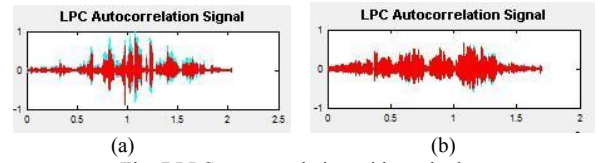


Fig. 7. LPC autocorrelation with method

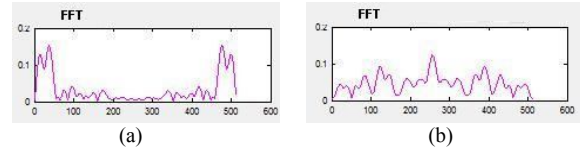


Fig. 8. FFT signal from LPC coefficients

V. CONCLUSION

According to the result and working through the research, it can be concluded that :

1. It can be designed in an autistic child's voice identification system based backpropagation neural network with MATLAB programming, in this case using the LPC extraction with autocorrelation method to extract the data of voice recording, followed by FFT process to sharpen information that used as input of ANN.
2. The most optimal in ANN training for this research used the number of neurons in layer hidden was 5, 300 epochs, and learning rate of 0.1.
3. Voice detection autistic children using artificial neural network, by utilizing the backpropagation method can be used as a diagnostic tool with the accuracy of the normal children voice was 100%, and 100% for the children voice with autism.

REFERENCES

- [1] 112.000 Anak Indonesia Diperkirakan Menyandang Autisme, Republika Online, Accessed on 2 Juli 2013
- [2] K. Oller, Warren, dkk, Automated Vocal Analysis of Naturalistic Recordings from Children with Autism, Language Delay, and Typical Development. PNAS, USA, 2010
- [3] Kaplan and Saddock, Kaplan and Sadock's Synopsis of Psychiatry, USA, 1994
- [4] Budiman, Melly, Makalah Simposium Pentingnya Diagnosa Dini dan Penatalaksanaan Terpadu pada Autisme, Surabaya, 1998
- [5] Santoso, Budi, dkk, Modul Pengolahan Sinyal Digital, Politeknik Elektronika Negeri Surabaya, Surabaya, 2012
- [6] David, Frederikus, Penggunaan Prosesor Sinyal Digital Keluarga TMS320 Sebagai Alat Pengenalan Suara Manusia Dengan Algoritma DTW (Dynamic Time Warping), UK Petra, Surabaya, 1996.
- [7] Rabiner, L., Biing-Hwang Juang, Fundamentals Of Speech Recognition, New Jersey: Prentice Hall, 1993
- [8] Kusumadewi, Sri, Membangun Jaringan Saraf Tiruan Menggunakan MATLAB dan EXCEL LINK, Penerbit Graha Ilmu, Yogyakarta, 2004